Here's some explanation about the files:

**IDBGenesV186.24 (explants)**

This is how the file was generated:

1) Analyze the microarray data using the R packages called lumi and limma

Limma should be cited this way:

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK (2015). “limma powers differential expression analyses for RNA- sequencing and microarray studies.” Nucleic Acids Research, 43(7), pp. e47.

Lumi should be cited this way:

Du, P., Kibbe, W.A., Lin and S.M. (2008). “lumi: a pipeline for processing Illumina microarray.” Bioinformatics.

2) Find probes that are differentially expressed with the conditions that p value < 0.05 and absolute values of the log fold change > or = 0.5

3) Enter those significant probes into InnateDB (http://www.innatedb.com/) in order to get the gene names associated with the probes and other information about them. I accessed InnateDB on 8Jan2016.

4) The raw output from innateDB is less easy to read, so I did some formatting in R to generate a file that would be easier to use. However, I can send you the raw file too if you want.

Here is what each column name means:

Treatment: This is either V186.24UP or DOWN

entrez: This is the entrez ID that corresponds to the Illumina probe. There may be multiple GO\_categories and GO\_terms per gene, which is why the genes are repeated several times.

name: This is the gene abbreviation associated with the entrez ID.

fullname: The full name of the gene

cerebralLocation: This is the location in the cell where you would find the product of that gene. It is called "cerebralLocation" because it is referring to the program called "Cerebral" that you can access through innateDB. I have looked at Cerebral a little bit, but not much.

GO\_category, GO\_term, GO\_Id: These refer to where this gene fits into the Gene Ontology. You can read more about GO (stands for "Gene Ontology") and what it means here: http://geneontology.org/

Fold Change: This is the log2 fold change of treatment over control at 24hrs.

Regarding missing genes from the file: There were some probes assigned to genes that do not yet have a GO\_Id assigned to them because the function is not known. I did not include those in the file but I can give them to you. I know that the genes you mentioned are well characterized, so if they are not on the list, it means that they were not among the genes that were not differentially expressed at the designated significance level for this treatment. Did you see them on another list related to this treatment or was there another reason that you expected them to be there?

**pathwayV186.24ORA (explants)**

This file was generated in the same way as IDBGenesV186.24 up to step 3. After getting entering the data into InnateDB, I did a Pathway Over-representation Analysis. This tests to see if any gene pathways are significantly associated with the genes that we already found to significantly up or down regulated.

Statistical method: I chose the hypergeometric test

Correction for multiple testing: Benjamini-Hochberg algorithm

I didn't do any manipulation of the results in R this time, I simply made one tab for the results that I got from plugging in the significantly UPregulated probes and another tab for the results from plugging in the significantly DOWNregulated probes, so this is what the raw data looks like when it comes out of innateDB.

Regarding your question:

There's a column that's called "Genes

(Symbol|IDBG-ID|Ensembl|Entrez|Fold Change|P-Value)" but I think the fold change and the P-value are missing. Do you know if I'm right with my guess?

I agree that there does not appear to be a fold change or P-value present in that column. I have noticed that unfortunately, innateDB column names are sometimes confusing and don't always correspond to the actual contents of the column very well. I wrote to the people in charge of InnateDB and they said that it is a mistake. Here is the email thread:

*Hi InnateDB folks,*

*I have a question about the results that I got from doing a pathway ORA. When I downloaded the results of the analysis, the resulting file has a column called “Genes (Symbol|IDBG-ID|Ensembl|Entrez|Fold Change|P-Value)”, but that column doesn’t seem to have any Fold Changes or P-Values in it, just all the different identifiers. Do you have any ideas about why these data aren’t present? Thanks a lot for your help!*

*Claire*

*Dear Claire,*

*Thank you for reporting this issue. I have confirmed it is a bug. This is a feature that used to be available but must have gotten lost in an update at some stage during development.*

*We are investigating to see if it is easy to add back in – but obviously this is data associated with the genes you uploaded so you should be able to link back to it relatively easily.*

*Kind regards,*

*David.*